Get ready to use EDirect

Step-by-step tutorial for Windows operating systems

Entrez Direct (EDirect) is a software package that provides access to NCBI's interconnected databases (publication, sequence, structure, gene, variation, expression, etc.). **EDirect** enables searching, linking to related records, retrieval of information, and extraction of specific data elements. You can easily combine individual operations in multi-step workflows to create custom datasets and generate reports.

> **i** **EDirect** is designed to run in a **Unix** command line environment and requires the **Perl** language interpreter to run its commands.
>
> **Unix** command line environment: **Unix** has been around a long time and is still a good operating system for issuing instructions to a computer by typing commands on a line. Although we now have more modern ways of interacting with a computer (graphical user interface with the ability to do mouse clicks or touch the screen), using a command line interface (in a terminal window) is still a powerful and flexible way of telling a computer what to do.
>
> **Perl** programming language: **Perl**, developed in the late 1980s and often referred to as the Swiss Army chainsaw of scripting languages, was adopted early and often as a common language for bioinformatics as it is powerful, accessible, and adaptable. **EDirect** uses some Perl commands, so to install and use **EDirect**, you will make sure that the Perl programming language is installed on your system.

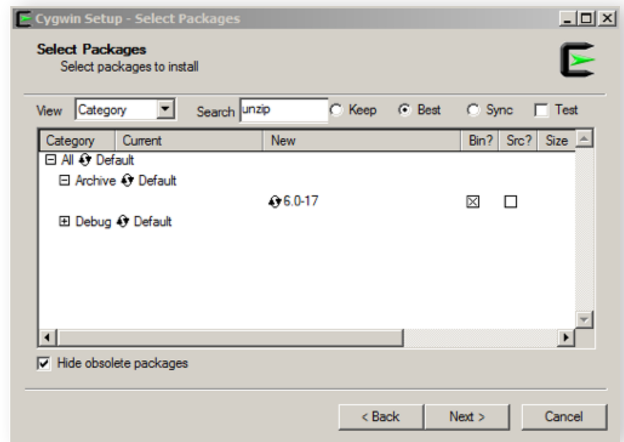## ① Install Cygwin helper application

Windows users need to install software that mimics a Unix command line environment and includes a Perl language package. **Cygwin**, a free, open-source application that allows you to interact with a Unix command line terminal, is a reliable application that offers a variety of add-ins that you can include in your installation, including programming languages like Perl and Python. It also enables quick installation of small utility add-ons including **Unzip** (required for installation of EDirect) and **Nano** (a simple text editor that you will use in this workshop).

1. Go to Cygwin's website (https://cygwin.com/install.html).

2. Begin the installation by clicking the appropriate version of Cygwin. (Select *"Run"* even though *"The publisher could not be verified."*)

3. Select "Install from Internet", leave the default "Direct Connection" selected, and click *"Next"*.

4. To choose your Download site, select the first entry: *"Cygwin.mirror.constant.com"* and click *"Next."*

5. To get around potential computer security issues with this installation, change the install directory from (C:\cygwin64) to a Cygwin directory on your Desktop (for example, C:\Users\your_user_name\Desktop\Cygwin). If the folder does not exist, it will be created. Select the *"Just me"* button and click *"Next"*.

Video of this tutorial at
https://youtu.be/hTzqsb1eD10

Select the Perl Package and Add-ons in the installation

EDirect requires the complete Perl package and the Unzip utility. The Nano text editor is a helpful accessory package that will be used in this workshop.

6. On the *"Select Packages"* page of the installer, make sure that *"Category"* is selected in the *"View"* drop-down menu at the upper left, then click the [+] to expand the *"All"* tree and find the *"Perl"* category in the list. To install all Perl packages, click on the circle-with-arrows icon (just to the right of the word *"Perl"*) until the label next to it switches from *"Default"* to *"Install"*.

7. In the search bar, type *"unzip"*. There will be two selections, Archive and Debug. Click on the [+] before Archive and an entry for unzip will appear. If it says *"skip"*, click on *"skip."* A version number will appear (6.0-16) to indicate this will be installed.

8. In the search bar, delete *"unzip"* and type *"nano"*. The list will display Debug and Editors. Click on the [+] before Editors and an entry for nano will appear. If it says *"skip"*, click on *"skip"* and it should change to a version number (2.8.7-1) indicating that it will also be installed.

Finish the full installation

9. Click on *"Next"* in the lower right corner of the window. A message regarding *"Review and confirm changes"* will appear. Select *"Next"*. Then click *"Next"* to start the download & installation. (It may stall and can take a while, but just let it go...)

10. Once installation is complete, a prompt for creating icons is displayed. You'll probably want to put an icon on the Desktop. Then, select *"Finish."*

② Install EDirect

To install EDirect, first open your Cygwin terminal window. Then, go to the EDirect help document (http://bit.ly/EDirectHelpDoc) and scroll down to *"Getting Started > Installation"*. Copy and paste the text below into the Cygwin terminal window.

```
cd ~
/bin/bash
perl -MNet::FTP -e \
    '$ftp = new Net::FTP("ftp.ncbi.nlm.nih.gov", Passive => 1);
    $ftp->login; $ftp->binary;
    $ftp->get("/entrez/entrezdirect/edirect.tar.gz");'
gunzip -c edirect.tar.gz | tar xf -
rm edirect.tar.gz
builtin exit
export PATH=$PATH:$HOME/edirect >& /dev/null || setenv PATH "${PATH}:$HOME/edirect"
./edirect/setup.sh
```

If you see the following message:

```
In order to complete the configuration process, please execute the following:
echo "export PATH=\$PATH:\$HOME/edirect" >> $HOME/.bash_profile
```

Copy and paste the line shown above (beginning with echo) into your terminal window.

Once the installation is complete, you may the following message in your terminal window:

```
Entrez Direct has been successfully downloaded and installed.
```

③ Test your EDirect installation

To confirm that EDirect is installed and working properly, run your first EDirect script!

Type the following into your terminal window and press Enter.

```
esearch -help
```

You should see this printed on the screen:

```
esearch 8.30
Query Specification
    -db       Database name
    -query    Query string
......
```

The 8.30 version shown on your screen is likely different as this program updates regularly.

Then type the following into your terminal window and press Enter.

```
xtract -help
```

You should see this printed on the screen:

```
xtract 8.30
Overview
  Xtract uses command-line arguments to convert XML data into a tab-delimited table.
  -pattern places the data from individual records into separate rows.
  -element extracts values from specified fields into separate columns.
  -group, -block, and -subset limit element exploration to selected XML subregions.

Processing Flags
  -strict           Remove HTML and MathML tags
  -mixed            Allow PubMed mixed content (Please note, the 8.30 version shown on your screen
......
```

You have now installed EDirect successfully and can begin using it!

Questions or feedback?
info@ncbi.nlm.nih.gov